BBC RD 1995/8

BBC
RESOURCES

Research and Development Report

LOW-DELAY PREDICTIVE AUDIO CODING FOR THE HIVITS HDTV CODEC

A.K. McParland, B.Sc., *and* N.H.C. Gilchrist, B.Sc., C.Eng., M.I.E.E.

Research and Development Department
Technical Resources Division
THE BRITISH BROADCASTING CORPORATION

BBC RD 1995/8
ISSN 1361-2379

LOW-DELAY PREDICTIVE AUDIO CODING FOR THE HIVITS HDTV CODEC

A.K. McParland, B.Sc. ***and*** **N.H.C. Gilchrist, B.Sc., C. Eng., M.I.E.E.**

Summary

This Report describes the work on predictive audio coding carried out by the BBC as part of the RACE 'HIVITS' project.

The purpose of the work was to develop and demonstrate a low bit-rate audio codec with low delay for use with a digital HDTV (high definition television) signal coded by the HIVITS HDTV codec and capable of conveying at least 5 channels of audio within the HDTV bitstream.

Issued under the Authority of

Research & Development Department
Technical Resources
BRITISH BROADCASTING CORPORATION

General Manager
Research & Development Department

(R017)

1995

LOW-DELAY PREDICTIVE AUDIO CODING FOR THE HIVITS HDTV CODEC

A.K. McParland, B.Sc. *and* N.H.C. Gilchrist, B.Sc., C. Eng., M.I.E.E.

LOW-DELAY PREDICTIVE AUDIO CODING FOR THE HIVITS HDTV CODEC

A.K. McParland, B.Sc. *and* **N.H.C. Gilchrist, B.Sc., C. Eng., M.I.E.E.**

1. INTRODUCTION

In the RACE* HIVITS** project the BBC worked in a European partnership to develop a low bit-rate digital television codec, intended primarily for HDTV contribution use at a bit rate of about 140 Mbit/s. The BBC worked with Thomson-CSF/LER and TRT, a French subsidiary of Philips, on the video coding, and with Deutsche Thomson-Brandt and CCETT on the audio coding. This Report decribes the BBC's work on audio coding for HIVITS completed in 1993; an earlier Report[1] describes the work on the video coding.

The work of the audio sub-project fell into two areas. CCETT and Deutsche Thomson-Brandt were particularly interested in using frequency-domain coding techniques to exploit not only the redundancy of the information contained in audio signals, but also the psychoacoustic properties of human hearing to maximise the bit-rate reduction obtainable.[2—5] The BBC concentrated on the addition of predictive techniques to the proven NICAM*** audio coding system,[6] to enable it to operate at lower bit-rates. The potential for reducing the bit-rate is much greater with frequency-domain coding than with NICAM, but a data capacity of 2.048 Mbit/s was available for audio and ancillary data in the HDTV contribution multiplex, and even a relatively modest degree of bit-rate reduction would be able to provide the necessary number of high-quality audio channels at this bit-rate. At the workshop, which was held at the completion of the HIVITS project, the BBC provided the five-channel audio coding and decoding for the HDTV contribution codec. Subsequently, the HDTV codec was demonstrated to members of the European Parliament, again using the BBC's five-channel predictive audio coding and decoding.

There were two reasons for deciding to develop a predictive codec based upon NICAM. The first was that is possible to make a codec with a relatively low delay using NICAM, as long as the coding block length is kept short. A long delay in the audio codec might make it necessary to place a compensating delay in the video codec, to keep the video and audio co-timed. This could prove expensive, but in any case excessive delays in contribution connections can often prove to be a nuisance.[7] Frequency-domain coding uses transform or sub-band techniques, and these have been found to introduce relatively long delays, sometimes more than 100 ms.[8] The second reason was that the need to provide some coding margin in audio contribution connections, to allow for subsequent processing of the audio signals, was recognised. At the start of the project, it was not known how quickly an audio contribution signal would deteriorate when subjected to various types of processing, or further coding and decoding operations. Some experience had, however, been obtained with NICAM. Subsequently, some tests were conducted on low bit-rate audio codecs, singly and with multiple passes of the audio through the codecs under test,[8, 9] and so some guidance is now available on the extent to which bit rate may be saved using frequency-domain coding in contribution connections.

The audio coding used an initial resolution of 16 bits/sample (i.e. 2 bits/sample more than conventional NICAM) with a sampling frequency of 32 kHz. The NICAM compression was effected in blocks of 32 samples, in each channel. Initial development was carried out using a Sun Workstation, which required digital audio signals to be recorded on the workstation's hard disk, processed and then replayed from the disk. Only the record and replay operations were in real time; the processing of the signals by the workstation was very much slower than real time. Subsequent development moved to a real-time processing system, which speeded up the work in many respects and also formed the basis of the coding and decoding hardware which was to become part of the HIVITS Demonstrator.

2. THE PREDICTIVE CODING ALGORITHM

2.1 Predictive coding

The coding algorithm used is essentially a type of ADPCM (Adaptive Differential Pulse Code Modulation). Straightforward DPCM consists of sending the difference between the current and previous audio samples. The previous sample can thus be viewed as an approximation to the current sample. An adaptive approach improves on this by creating (hopefully) a

* **R**esearch into **A**dvanced **C**ommunications in **E**urope.

** High Quality Video Telephone and HD(TV) Systems.

*** **N**ear-**I**nstantaneous **C**ompanded **A**udio **M**ultiplex.

better approximation to the current sample.

One way to obtain a better approximation is to use prediction. This involves linearly combining previous sample values to generate an estimate of the next value. The prediction will be good if the signal contains redundancy, which can be exploited. If the prediction is good, then the number of bits needed to code the difference signal will be relatively small. Thus the bit rate may be reduced, but no information is lost, so perfect reconstruction can take place at the receiver. In other words, the additional saving in bit rate is obtained by discarding redundant information.

A functional diagram of a predictive coder is shown in Fig 1. The audio samples applied to the coder (at the extreme left-hand side of the diagram) firstly undergo a prediction process. This involves subtraction of the predicted value from each sample, such that the output is a prediction error signal. A number of predictors are available, the one giving the lowest r.m.s. (root mean square) prediction error for a block of 32 samples is selected for that block. The block length of 32 (samples) was chosen to correspond with that of the NICAM process (see below). The coder sends the decoder a short control word with each block to identify the predictor in use.

A requantiser is needed because the prediction will vary in its accuracy, and there is a need to regulate the bit-rate. This is done by providing a fixed number of bits per sample for coding the prediction error signal. Small prediction error signals are coded with a high resolution; larger prediction error signals, which occur when the prediction is less good, are coded with a lower resolution, thus losing some information. Two processes were tried for the non-linear requantisation of the prediction error signal: A-law and NICAM. A-law is an instantaneous companding system based upon a non-linear segmented characteristic; NICAM is a block companding system.[10] NICAM was found to introduce much less coding error than A-law.

In NICAM, the highest amplitude sample in a block determines which bits of each sample are transmitted. When NICAM is used in a differential coding arrangement, it is the largest difference-signal sample in each block which determines the bits which are transmitted. Only the most significant bit (sometimes termed the 'sign bit') and the most significant of the remaining active bits are selected for transmission. If all of the samples in a block are at a low level, then only a few least significant bits will be active. When the difference signal is at a higher level, there are too many active bits to be transmitted, and some of the least significant bits have to be omitted from the coded signal. The 'lost' least significant bits represent lost information (i.e. a reduction in the resolution) but the most significant bits which are not transmitted can be regenerated at the decoder from the sign bit. Thus, only low-level difference signals will be quantised with the full resolution. The bits that are transmitted are identified for the decoder by a scale factor conveyed with the audio difference samples. Another term for this is floating-point block companding, the requantised samples being the mantissae and the scale factor the exponent. The standard NICAM block comprises 32 samples (i.e. 1 ms at 32 kHz sampling frequency). Error feedback is used to reduce the effects of requantisation. The block 'Z' in Fig. 1 could be a simple delay or a filter.

2.2 The choice of predictors

The chosen set of predictors must cover the complete range of input signals. One way to create a predictor is to 'train' an adaptive predictor on the signal to be coded. General purpose predictors can be made by

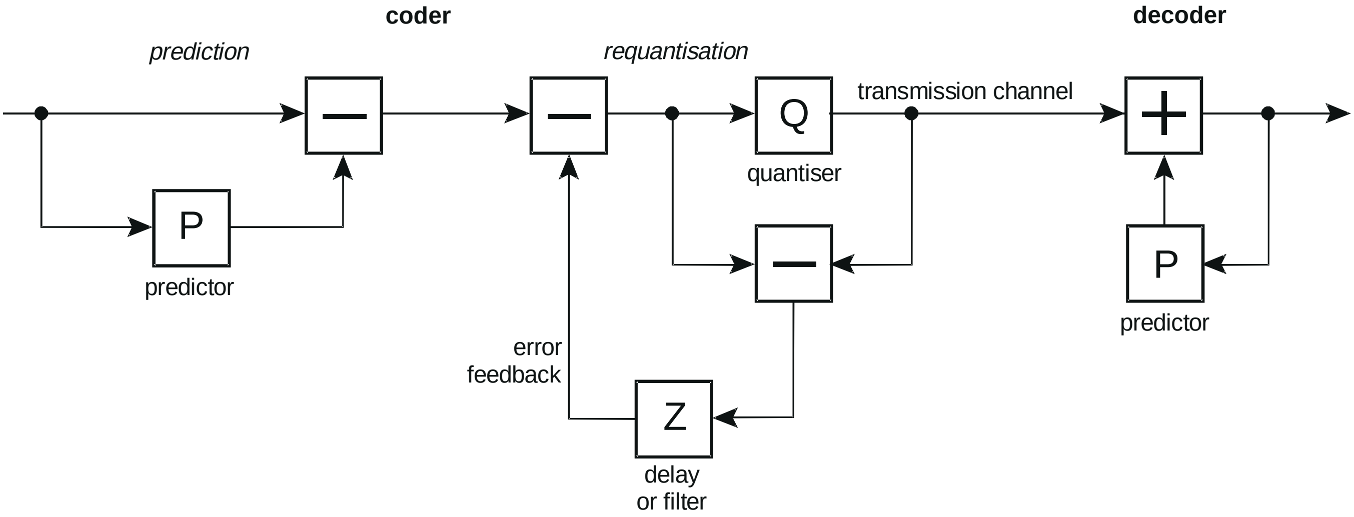

Fig. 1 - Functional diagram of the predictive coder and decoder.

using a wide range of material. If a specifically critical extract is used during the optimisation of a predictor, the result will be a predictor tuned to that particular signal. A combination of general purpose and specific predictors is needed for a general purpose coding system.

When designing a predictor, it is necessary to determine the type of audio material for which it should be optimised. The predictor coefficients are obtained by finding values which minimise the prediction error.

Predictors can be considered as filters. The coding ‘filter’ will ideally produce an error signal which resembles white noise. The decoding ‘filter’ will reconstruct the spectrum of the original signal. The frequency response of the predictor in the decoder should therefore match the spectrum of the input signal. This leads to another method of predictor design. The poles of the all-pole decoder ‘filter’ can be specified and then the predictor coefficients worked out. The frequency responses of the various predictors can be plotted and compared. The predictors in the coder are *finite impulse response* (FIR) filters which operate upon the input samples; but in the decoder, the predictors take the form of *recursive* or *infinite impulse response* (IIR) filters. These operate on requantised prediction error values to reconstruct the digital audio samples. Recursive filters can become unstable, if care is not taken in selecting the coefficients. The instability may become evident if bit errors occur in the channel between the coder and the decoder, or as the result of requantisation; this instability manifests itself as oscillation in the decoder. It is most important, therefore, that only stable predictors are used in the codec.

Some study was devoted to the possibility of using adaptive predictors. The principle of the adaptive predictor is to adopt a general form of predictor; but to choose values for the coefficients which minimise the prediction error for each 32-sample block. In other words, the predictor is optimised for each block by adjusting the coefficient values. When this was tried, two problems were encountered. Firstly, there were difficulties in keeping predictors using coefficients calculated on a block-by-block basis stable; sometimes the audio signals reconstructed in this way were very distorted. Secondly, the improvement obtained when the adaptive prediction was stable was generally not sufficient to justify the additional bit-rate needed to signal the coefficient values to the decoder.

In the final implementation of the codec, fixed predictors were used with lengths of up to 4 taps (i.e. 4 previous samples). With these lengths, the prediction cannot achieve very good signal-matching characteristics, and the difference signal cannot be adequately conveyed with a very low number of bits. Longer predictors give better results, but they are then more signal-specific and take more processing time to implement. Furthermore, because the predictors are now more closely attuned to the requirements of specific signals, more predictors need to be provided to ensure good performance with a wide range of programme material. This in turn means that there are more predictors to be tested with each block of samples, placing heavier demands on the processing in the coder. So with practical limitations on the processing power which can be provided, a compromise has to be made between performance and generality when deciding which predictors to use in the coder.

2.3 Preliminary software development

The initial software development was performed in ‘C’ on a SUN Unix workstation in non-real-time. Test audio programme items were captured using an AES/EBU I/O (input/output) interface to the SUN, developed by the BBC. This interface was a VME bus card which plugged into the VME backplane of the SUN workstation. The resultant VME-based system gave real-time recording and playback from the SUN’s hard disks. The test items were processed and played back via the same interface. The workstation environment meant that the audio data could be viewed and analysed easily, but the processing was at a speed about 120 times slower than real-time.

Some listening tests were carried out to confirm the viability of the coding system. Several critical programme items were processed with different sample word lengths. Good results were obtained at 7 bits per coded sample, approaching NICAM quality, but 8 bits were needed to give very low levels of impairments.

With the acquisition of general-purpose Digital Signal Processor (DSP) hardware, the algorithm could be tested in real-time. The limitations on the speed of the processor meant that not as many predictors could be tried, but testing and software development could proceed more rapidly.

3. THE IMPLEMENTATION OF AN EXPERIMENTAL SIX-CHANNEL CODEC

A six-channel audio codec was constructed to serve both as a development tool and as an audio component of the Demonstrator for the HIVITS project. A brief specification of the codec is given in the Appendix to this Report.

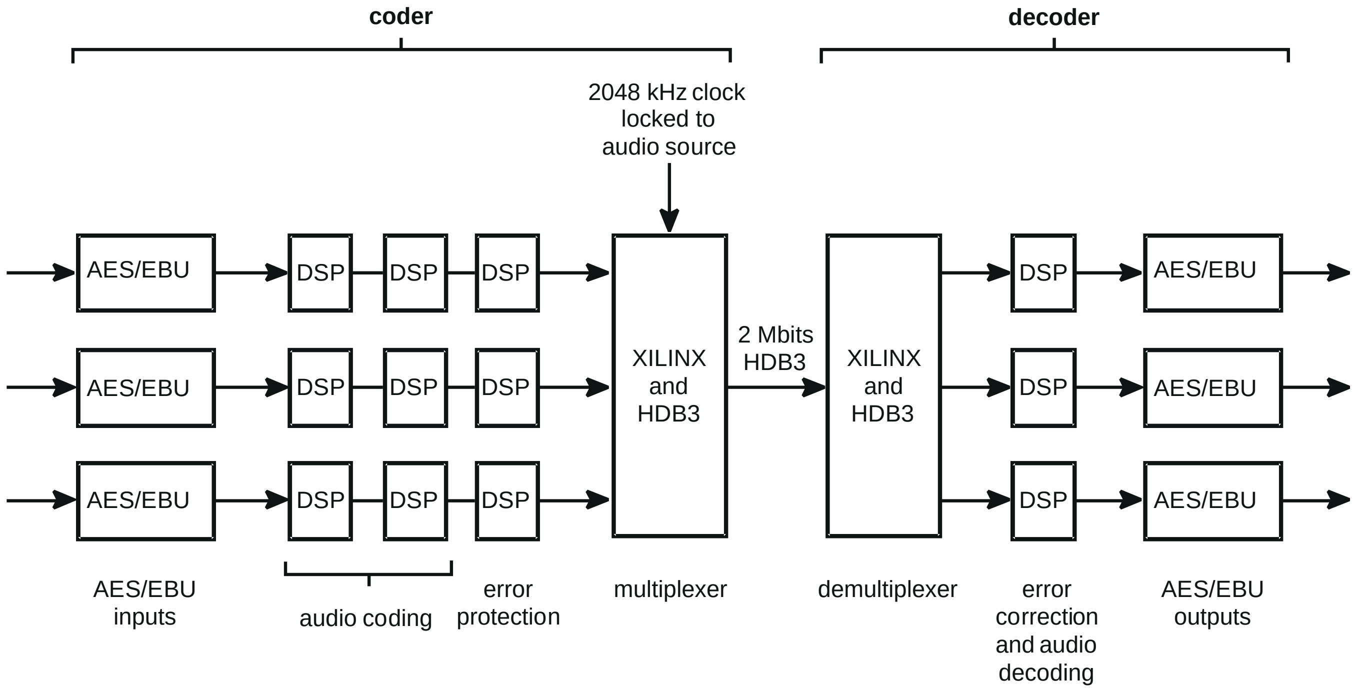

Fig. 2 - Block diagram of 6-channel codec.

3.1 Hardware

The real-time hardware for the coder and decoder is based on the AT&T DSP32C floating-point digital signal processor (DSP). The choice of a floating-point processor with a good ‘C’ compiler facilitated the simple transfer of the coding algorithm. Each processor unit comprises a 3U Eurocard containing a DSP, 256 kbytes of zero wait-state RAM and a micro-controller to facilitate control and downloading of software. ‘C’ programs facilitate spreading of the processing load by chaining together processor cards using a fast serial interface. The processor cards are carried in a rack, equipped with power supplies. AES/EBU input and output units are also contained in the rack.*

Software development for the real-time hardware involved initially transferring the existing ‘C’ code to the DSP. The ‘C’ compiler was reasonably effective, but the most processor-intensive tasks were rewritten in assembly language.

A six-channel analogue-to-digital converter rack was constructed for this project to handle analogue inputs. It is based on previously-designed ADC cards, with BBC AESIC AES/EBU outputs, and runs at a sampling frequency of 32 kHz. This rack also provides the reference 2.048 MHz clock for the multiplex, and ensures all the AES/EBU sources are locked together and to the coder clock.

Fig. 2 shows a block diagram of the 6-channel codec. Three DSPs are used for each pair of channels in the Coder: 1 each for predictively coding the ‘A’ and ‘B’ channels, and 1 for multiplex preparation/error protection. One DSP is used per channel-pair in the Decoder.

The multiplexer and demultiplexer cards comprise a XILINX field-programmable gate array (FPGA), an EPROM to program it and HDB3 code-conversion circuitry; the demultiplexer contains, in addition, a phase-locked loop. In the coder, the XILINX design generates its own frame-alignment word and gets 32-bit words from each of the error-protection DSPs using the serial Direct Memory Access (DMA) feature of the DSPs.

3.2 Reducing the number of predictors

In the initial simulations and tests, over 50 predictors were available to the coder. This is too many for a real-time design with a realistic number of DSPs. Some simulation work was carried out to determine which predictors were the most frequently selected and which produced the lowest coding errors. The most useful predictors were selected for the software to run in the real-time hardware.

With this hardware, the predictors could be assessed subjectively with many items of material and

* The processor units, rack and AES/EBU interface units were suppied by the Fraunhofer Institute for Integrated Circuits, Erlangen, Germany.

predictors removed to see if their removal impaired the signal significantly. Reducing the number of bits per sample simplified the decision-making process, by increasing the prediction error signal. This was necessary, as only about 8 predictors could realistically be evaluated in the time available for the work.

3.3 Multiplex

The multiplex in this implementation generates a 32-bit frame-alignment word for a 2048-bit frame structure (i.e. a frame period of 1 ms). The data from each pair of channels is then inserted into the frame in turn. This data consists of processor synchronisation words followed by predictor, scale factor and audio data, with associated error protection words.

In this implementation, 8 bits per sample are used to code the audio, giving a net bit-rate of 264 kbit/s per audio channel (see Appendix), as there is sufficient capacity. With different protection, 7 channels could be fitted into the bitstream. CCITT Recommendation G.704 framing could be used with 6 channels if there is a requirement for this.

A simple error protection scheme can be employed. This uses a single burst-error correcting cyclic code, (151,136), which will correct any bursts up to 6-bits in length. The predictor number and scale-factor are repeated within the protection of the code. Checks are made on the validity of the data to aid in error concealment.

4. OPERATING THE EXPERIMENTAL AUDIO CODEC IN CONJUNCTION WITH A RACE 1018 VIDEO CODEC

The first opportunity to use the audio codec in conjunction with the video codec occurred at the HIVITS Open Days, held at BBC Research Department, in May 1992. At this time, the HDB3

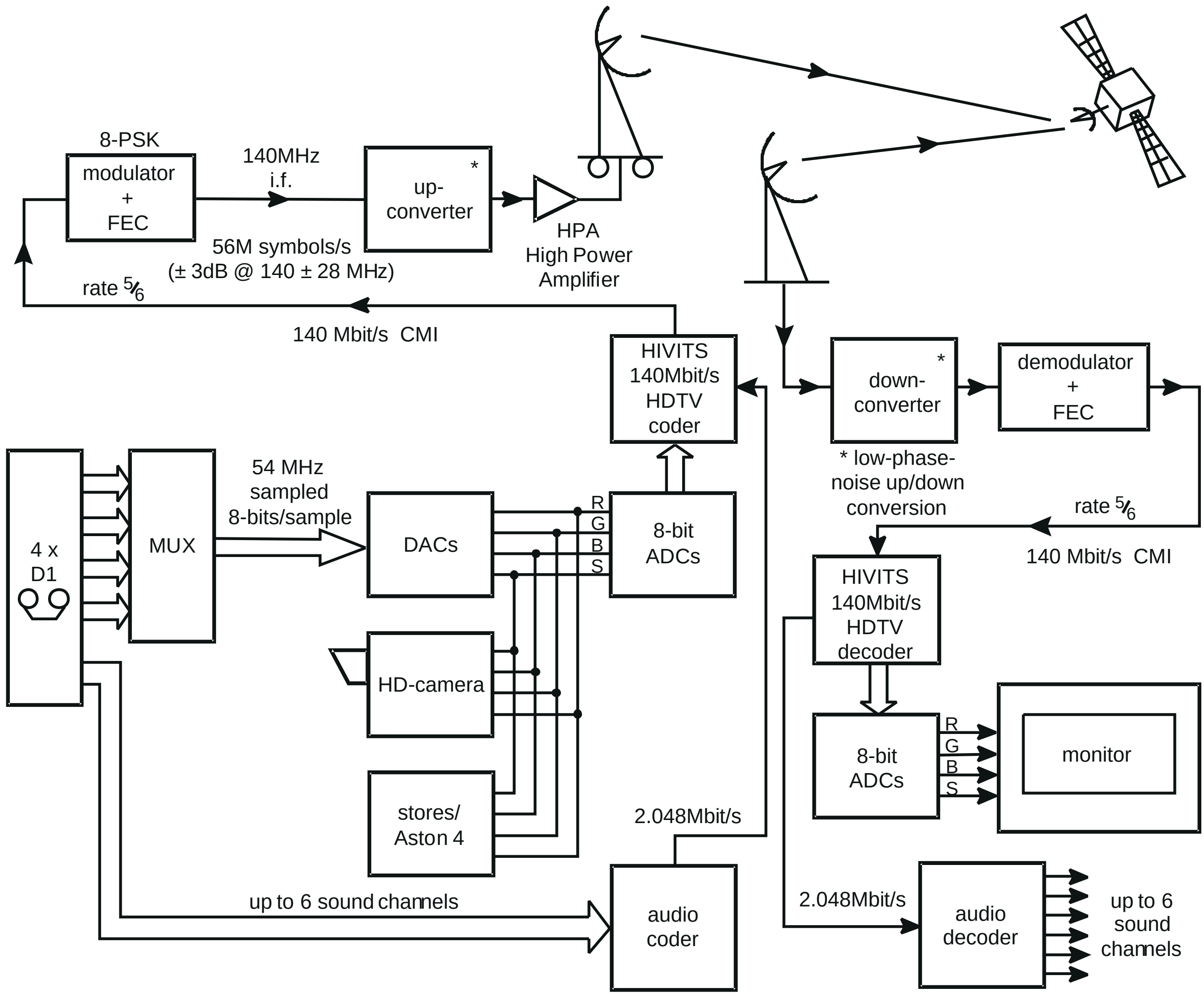

Fig. 3 - Block diagram showing the equipment used to provide an HDTV ‘contribution’ quality link via satellite.

interface in the video codec was not functioning, so the audio coder and decoder were demonstrated connected directly to each other, while the video coder and decoder were connected via a satellite link.

Once the HDB3 interface was working, the audio could be sent in the multiplex with the video. There were severe timing jitter problems with the decoded HDB3 audio signal, so a well-damped phase-locked loop could not be used. The phase-locked loop had to track the jitter. The error performance of the audio channels were checked by introducing errors into the transmission channel. With random and burst errors, the audio failed at a slightly lower bit-error ratio than the video.

The next major demonstration of the two codecs working together was the HIVITS Workshop held at Thomson LER in Rennes, France on 19th-20th January, 1993. Here the coders were placed in a studio, and an optical fibre link was used to connect the coders to the decoders, which were situated in a remote building. This demonstration was the culmination of the project. Continuous demonstrations were run, and there was a papers session where all aspects of the project were presented.

A further major demonstration of the system was given to the European Parliament in Brussels during 9th-11th June 1993. This comprised a demonstration of bit-rate-reduced HDTV and multi-channel sound sent via satellite from London, using the HIVITS video and audio codecs. Fig. 3 *(see previous page)* shows a simplified block diagram of a HIVITS HDTV (with multichanel sound) contribution-quality connection via satellite.

5. CONCLUSIONS

A predictive audio coding algorithm has been developed, as part of the RACE HIVITS project, and six-channel audio coding and decoding hardware developed. Though the compression factor is relatively modest, an advantage of predictive coding over some of the more advanced techniques is that it can be implemented with relatively low signal delay; in this case, the measured delay through the coder and decoder was 7 ms.

A number of demonstrations of the audio codec have been given, operating in conjunction with the video codec to provide an HDTV contribution connection.

6. ACKNOWLEDGEMENTS

The authors would like to acknowledge the contribution of the BBC's Partners in this RACE project, and to thank Dr. E.F. Schröder of Deutsche Thomson-Brandt and Mr. Y.-F. Dehery of the CCETT for helpful discussions during the course of this work. The work described was part of RACE project No. 1018 which was supported by the Commission of the European Community through the Framework Programme.

7. REFERENCES

1. COTTON, A.D.R. *and* WELLS, N.D., 1992. Experiments with the HIVITS HDTV contribution codec. BBC Research Department Report No. 1992/12.

2. DEHERY, Y.-F., 1991. MUSICAM source coding. Proceedings of the 10th AES International Conference — Images of Audio, (London) pp.71-79.

3. WIESE, D. *and* STOLL, G., 1990. Bit-rate reduction of high quality audio signals by modelling the ear's masking thresholds. AES 89th Convention (Los Angeles). AES preprint No. 2970.

4. JOHNSTON, J.D., 1988. Transform coding of audio signals using perceptual noise criteria. IEEE Journal on Selected Areas in Communications. **6**(2) (February), pp. 314-323.

5. MOORE, B.C.J., 1993. Characterization of simultaneous, forward and backward masking. Proceedings of the 12th AES International Conference — The Perception of Reproduced Sound, (Copnhagen) pp. 22-33.

6. CAINE, C.R., ENGLISH, A.R. *and* ROBINSON, J., 1980. NICAM 3: A companded PCM system for the transmission of high quality sound programmes. Proceedings of the International Broadcasting Convention (Brighton). IEE Conference Publication No. 191, pp. 274-277.

7. GILCHRIST, N.H.C, 1991. Delay in broadcasting operations. AES 90th Convention (Paris, 1991). AES Preprint No. 3033.

8. GILCHRIST, N.H.C., 1993. Digital sound: The selection of critical programme material and preparation of recordings for CCIR tests on low bit-rate codecs. BBC Research Department Report No. 1993/1.

9. THIBAULT, L., GRUSEC, T., LYMAN, S. *and* DIMINO, G., 1994. Audio quality in digital broadcast systems. Proceedings of the Second International Symposium on Digital Audio Broadcasting (Toronto), **1**, pp. 405-420.

10. MCNALLY, G.W. *and* GILCHRIST, N.H.C., 1979. The use of a programmed computer to compare the performance of digital companding systems for high-quality sound signals. EBU Review-Technical, No. 178 (December).

APPENDIX

Summary Specification of the Audio Coding/Decoding Hardware

Bit-rate of multiplex:	2048 kbits/sec
Interface coding:	HDB3
No of channels:	6
Sampling-rate:	32 kHz
Bits per sample, input:	16
Bits per sample, coded:	8
Coding system:	block-based adaptive prediction using fixed list of predictors with NICAM requantisation
Block size:	32 samples (1 ms)
No of predictors:	8
Bit-rate per audio channel, including sync & error protection:	336 kbits/sec
Net audio bit-rate per channel:	264 kbits/sec
Ancillary data bit-rate:	12 kbits/sec
Inherent coding delay:	1 ms
Overall (measured) delay:	approximately 7 ms (most of this results from the chaining of processors)
Error protection:	(151,136) 6-bit Single-Burst Error Correcting Code
Number of DSPs:	12 (9 in coder, 3 in decoder)
Type of DSP:	AT&T DSP32C

Published by BBC RESEARCH & DEVELOPMENT DEPARTMENT, Kingswood Warren, Tadworth, Surrey, KT20 6NP **ISSN 1361-2379**